УДК 025.4.03

О.Р. Чертов, Д.В. Райчук
Национальный технический университет Украины «Киевский политехнический институт», г. Киев
chertov@i.ua

Адаптивная поддержка сотрудничества при поиске информации

В связи со значительным ростом объема информации, доступной в цифровой форме, и увеличением разнообразия веб-приложений, необходимость в эффективном информационном поиске становится критической. В статье предлагается применение технологии адаптивной поддержки сотрудничества, что обеспечит синергию работы пользователей поисковой системы.

Введение

Американский ученый Кельвин Мурс (Calvin Mooers) еще в 1948 г. в своей докторской диссертации ввел термин «информационный поиск» (information retrieval). Согласно [1, с. 2], под информационным поиском понимаются действия, методы и процедуры, позволяющие осуществлять отбор определенной информации из массива данных. Информационный поиск осуществляется при помощи информационно-поисковых систем (ИПС). Главная задача любой ИПС – быстрое формирование точного и полного ответа, адекватного запросу пользователя. Первые поисковые системы использовались преимущественно в библиотечном деле и в системах научно-технической информации, например, CDS/ISIS, DIALOG, STAIRS [2, с. 124]. Ситуация кардинально изменилась после широкого распространения Интернета:

− расширились возможности по опубликованию документов и удаленному доступу к ним, сам доступ стал значительно дешевле;

− круг как внешних, так и внутрикорпоративных пользователей ИПС, не имеющих специализированной подготовки, резко расширился, возросла специфика генерируемых ими запросов, уменьшилось время, которое типичный пользователь готов был выделить на поиск информации;

− отсутствие, как правило, при публикации в Интернете редакторского контроля породило дублирование информационных материалов, наличие некачественной информации (устаревшей, неправдивой, содержащей фактические и/или грамматические ошибки).

Следует также учесть, что по данным исследовательской компании IDC [3, с. 3] к 2011 г. общие объемы информации, хранимой во всем мире на всех существующих цифровых носителях, превысят $1{,}8\times10^{21}$ байт, а темпы ежегодного прироста этого объема составят почти 60%. В то же время, в большинстве ИПС поиск информации реализован на основе поиска по ключевым словам, который базируется не на ее семантике, а на некоторых вторичных свойствах (например, частоте встречаемости слов в документе). Этот механизм оказывается малоэффективным.

Одним из возможных путей повышения эффективности поиска является использование различных адаптационных механизмов, позволяющих существенно сузить его пространство. ИПС является адаптивной, если в ходе своей работы она подстраивается под конкретного пользователя. Применявшаяся еще в ряде первых ИПС, технология адаптивной фильтрации информации (adaptive information filtering) носит

вспомогательный характер. Она лишь позволяет упорядочивать и частично отфильтровывать нерелевантную информацию, исходя из заранее известных предпочтений пользователя, например, сервис My Yahoo! [4]. Только в 2008 г. некоторые ведущие глобальные поисковые системы начали предоставлять в экспериментальном режиме службы по «социализации» поиска, т.е. адаптивной поддержки сотрудничества группы пользователей при поиске информации: служба Википоиск (SearchWiki) [5] компании Google и служба U Rank [6] поисковой системы Live Search компании Microsoft.

Предлагаемые этими системами сервисы предполагают участие пользователей в формировании контента: пользователь может изменять структуру и порядок выдачи для себя найденной информации, оценивать результаты поиска. Делать окончательные выводы об эффективности новых возможностей такого поиска еще рано. Остаются открытыми вопросы об активности, компетентности и добросовестности пользователей, конфиденциальности их работы. И хотя рассмотренные системы не позволяют пользователям влиять на ранжирование общего поиска и используются пока в основном для изучения поведения пользователей, знаковым является уже то, что ведущие поисковые системы обратили свое внимание на аспект социализации поиска.

Вполне возможно, что рассмотренные проекты претерпят определенные изменения или будут даже приостановлены, как это произошло с проектом WikiaSearch (Wikipedia), однако есть все основания предполагать, что такие ИПС – это сервисы будущего.

В то же время, в ряде смежных областей, прежде всего, в области проектирования и использования автоматизированных обучающих систем, адаптационные технологии используются давно и успешно [7-9]. Поэтому перспективным представляется применение при поиске информации соответствующих наработок из области автоматизированного обучения. Тем более, что преимущества коллективного обучения (на основе сотрудничества) над индивидуальным известны еще со времен перипатетической школы Аристотеля.

Целью данной работы является разработка подхода к организации адаптации ИПС к ее пользователям, расширяющего возможности традиционного (для автоматизированного обучения) метода адаптивной поддержки сотрудничества путем формирования комплексной модели пользователя, в том числе на основе информации о других пользователях – представителях близких групп. При этом должно учитываться то, что со временем цели и интересы каждого пользователя могут не только конкретизироваться, но и изменяться.

Адаптация, адаптированные, адаптируемые и адаптивные системы

Термин «адаптация» произошел от латинского слова adaptatio, означающего «приспособление». Впервые в научный оборот понятие «адаптация» было введено в 1865 г. немецким физиологом Германом Аубертом (Hermann Aubert) для характеристики явления приспособления органов зрения к воздействию внешних раздражителей путем повышения или понижения порога чувствительности [10]. Понятие адаптации является полисемантичным и междисциплинарным. В технических науках и кибернетике под адаптацией понимают «накопление и использование информации для достижения оптимального, в некотором смысле, состояния или поведения системы при начальной неопределенности в изменяющихся внешних условиях» [11].

Система еще на этапе проектирования может быть адаптирована к конечному пользователю при условии, что известна специфика его поведения. Такая система называется адаптированной (adapted).

В процессе функционирования системы ее адаптация происходит путем целенаправленного изменения:
– либо параметров и структуры системы, тогда такая система называется адаптируемой (adaptable),
– либо алгоритма функционирования системы, тогда такая система называется адаптивной (adaptive).

«Адаптивность – это способность индивида или системы модифицировать себя или свое окружение, когда происходит неблагоприятное для них изменение для того, чтобы хотя бы частично восполнить потерю эффективности» [12, с. 129]. В отличие от адаптируемой системы, изменяющейся на этапе сопровождения с помощью внешних воздействий (в том числе, и человека), адаптивная система должна быть самонастраивающейся в реальном времени [13].

Целью адаптивных ИПС является предоставление пользователю помощи на каждом шаге информационного поиска, вплоть до предугадывания – выполнения следующего шага за пользователя. Для этого система должна проводить наблюдения за действиями пользователя, понимать (распознавать) их и использовать это понимание для предоставления помощи и корректирования (изменения) самой себя. Адаптация может использоваться как для полного контроля над поиском, так и для помощи в выборе наиболее подходящих документов.

Адаптация на основе коллективного взаимодействия

Контингент пользователей современной ИПС гетерогенен: различны знания людей, опыт работы, поведение, требования, запросы к системе. К тому же характеристики пользователя постоянно меняются во времени. Поэтому система в процессе адаптации к пользователю, возможно, будет находиться в неустойчивом состоянии. Постоянные изменения в системе, вероятно, могут привести к неприятию системы пользователем. Но если процесс адаптации будет достаточно быстро сходиться к оптимальному процессу, качество работы адаптивной системы может мало отличаться от оптимального. Следовательно, необходимо найти такой подход, при котором система достаточно быстро сможет перестраиваться в эффективную структуру, пригодную для производительной работы конкретного пользователя.

Ключевой особенностью применения адаптивных технологий является возможность проведения анализа деятельности большого числа пользователей одновременно. Результаты этого анализа могут использоваться при настройке функциональности для конкретного пользователя. Адаптация на основе коллективного взаимодействия значительно более привлекательнее адаптации на основе только личного опыта, так как в любой момент времени можно найти готовое решение, принятое на основе уже накопленного опыта.

Основные механизмы коллективного взаимодействия объясняются психологическими теориями, рассматривающими коллективную работу с трех позиций:
– социально-психологической (члены группы зависят друг от друга);
– когнитивно-психологической (развитие познавательных функций человека);
– бихевиористской (положительное подкрепление результативной работы группы).

Суть группы – во взаимозависимости ее членов, порожденной наличием общей цели. Жизнь группы – это динамическое целое, где изменение в состоянии одного из членов автоматически влечет изменения в состоянии всех других.

Общение пользователей может быть как явным (посредством электронной переписки, форумов, блогов), так и, скорее, неявным: пользователи отмечают полезные и бесполезные документы, оставляют какие-либо рекомендации и отзывы по поводу

конкретных документов. Коллективное взаимодействие может также выражаться в социальной навигации на основании данных о действиях других пользователей относительно текущей страницы (количества посещений страницы или времени, проведенного на ней другими пользователями). Подобная информация помогает понять, насколько данная страница релевантна сообществу пользователей. Для этих же целей могут применяться различные рейтинги или системы оценивания.

В результате, даже не зная друг друга, пользователи обмениваются между собой полезной информацией.

Перспективной представляется адаптация на основе коллективного взаимодействия активного пользователя с группами пользователей, имеющими близкие с ним предпочтения и цели [14]. Такая адаптация даже новичку дает возможность быстрого эффективного старта и профессионального поиска информации (на основе наработок своих коллег), сразу погружает его в профессиональный сленг и, как минимум, психологически позволяет ощутить поддержку со стороны более опытных товарищей.

Параметры модели пользователя

Исходя из целей построения адаптивной ИПС очевидно, что в ее основу должна быть положена модель (профиль) пользователя. При формировании такой модели нужно исходить из данных о знаниях, целях, предпочтениях индивидуального пользователя и информации о его взаимодействии с системой.

Модель пользователя должна состоять из следующих групп параметров:
– демографических;
– когнитивных и психологических;
– физиологических;
– профессиональных;
– параметров, характеризующих особенности взаимодействия с системой.

От продуманности и полноты параметров модели пользователя зависит эффективность процесса принятия решений по адаптации. Модель пользователя создается как из явно указанных, так и неявно полученных данных.

К явно указанным данным относятся:
– данные, которые сам пользователь указал при регистрации в системе (пол, возраст, интересы, предпочтения);
– рейтинги (оценки), выставленные пользователем найденным при поиске документам;
– явное указание того, какие инструменты (формы) стоит показывать или прятать.

Явно указанные данные полезны, так как являются прямым указанием системе, что именно интересно пользователю. Однако поддержка актуальности этих данных, как правило, редко обеспечивается пользователем, ибо требует от него выделения дополнительного времени. При создании модели пользователя явные данные учитываются, но им устанавливается меньший приоритет, чем неявным.

Неявно полученные данные – это те, которые ИПС формирует самостоятельно, без участия пользователя в ходе наблюдения за его действиями в системе. К неявно полученным данным относятся:
– данные моделей представителей групп, близких по предпочтениям: система выявляет, с какими группами данный пользователь имеет наибольшее сходство и предполагает, что решения, принятые представителями этих групп, ему приемлемы, – тем самым обеспечивается адаптивная поддержка сотрудничества; при этом данные из моделей «соседей» добавляются в модель текущего пользователя не навсегда: возможно, система ошиблась со своим прогнозом, и поэтому со временем придется найти другие подходящие группы для текущего пользователя;

– значимые слова (термы), которые содержатся в документе, прочитанном пользователем, если он провел за его чтением больше определенного времени: в работе [15] предложен метод конструирования модели пользователя в виде вектора, элементы которого являются частотами термов в отмеченных пользователем документах;
– приоритетность использования различных форм/инструментов, предлагаемых адаптивной системой: если пользователь редко использует какую-либо форму, то ее можно убрать, чтобы она не отвлекала внимание от более важных форм.

При формировании модели пользователя нужно учитывать как его долгосрочные предпочтения, так и предпочтения текущего сеанса работы. Как показало исследование [15], последние должны иметь наибольший приоритет.

Формирование групп пользователей. Для разбиения всего сообщества пользователей системы на группы целесообразно применять методы кластерного анализа. И хотя нет пользователей с абсолютно одинаковыми характеристиками (одинаковыми моделями пользователя), они могут быть «связаны» похожими характеристиками. Параметры, по которым будет проведено разбиение на группы, являются параметрами модели пользователя. Например, это могут быть знания, предпочтения, цели, уровень подготовки, имеющийся опыт пользователя, а также термы документов, которые он предпочел. Описание кластера (группы) пользователей – это значение параметров, на основе которых проводился кластерный анализ и получено текущее разбиение на группы.

При реализации механизма групповой адаптации ИПС будет принимать решения на основе модели текущего пользователя, данные в которой получены как непосредственно от самого пользователя, так и от представителей близких групп. Пользователь может быть проинформирован о том, значения каких именно характеристик получены от других пользователей. Если ИПС теряет эффективность поиска или пользователь пришел к выводу, что адаптация по указанным характеристикам не адекватна его цели, тогда система берет в работу следующую группу с меньшей мерой сходства с данным пользователем.

ИПС должна регулярно проводить переразбиение сообщества пользователей на группы, поскольку предпочтения и цели пользователей со временем могут меняться, а опыт по работе с системой – накапливаться.

Организация адаптивной поддержки сотрудничества при поиске информации

Ниже изложены основные шаги организации адаптивной поддержки сотрудничества пользователей при поиске ими информации.

1. При регистрации в системе пользователь имеет возможность заполнить данные о себе. Это могут быть как возраст, пол, образование, так и перечень интересов, сфер деятельности. Эти данные будут использованы для формирования первоначальной модели пользователя, необходимой для адаптации системы к его предпочтениям. Особого приоритета эти данные не имеют, так как не все пользователи их указывают, а тем более не все их поддерживают в актуальном состоянии. Однако для первоначальной адаптации системы они вполне приемлемы.

2. Система анализирует текущую модель пользователя, находит «близкие» группы пользователей, с которыми у данного пользователя имеется наибольшая мера привязанности (схожести). При этом учитываются приоритеты различных реквизитов.

3. Осуществляется адаптивная поддержка сотрудничества [8] путем дополнения модели пользователя характеристиками центроида наиболее близкой к нему группы (из списка всех, еще не перепробованных групп).

4. На основании исторических данных о работе в системе членов отобранной группы (или групп) подбирается (формируется) модель пользователя. По выбранной (созданной) модели однозначно подбирается пользовательский интерфейс и соответствующие методы его адаптации:
– переструктурирование гиперпространства для лучшей ориентации и передвижения в нем;
– показ справочного окна для помощи пользователю в достижении цели;
– предоставление дополнительных объяснений по некоторым понятиям;
– удаление элементов интерфейса, которые понижают эффективность пользователя или на нее не влияют;
– добавление новых элементов интерфейса, которые, возможно, помогут пользователю;
– изменение вида, стиля отображения показываемых форм.

5. Пользователь работает в рамках текущей модели: вводит поисковые запросы, просматривает документы, оставляет отзывы (оценки) найденным документам. При этом система отслеживает каждое действие пользователя в разрезе его информационно-поисковой работы. Новая информация заносится в модель текущего пользователя и получает наивысший приоритет.

6. Адаптивная система анализирует то, насколько эффективно производится поиск информации пользователем (каково соотношение найденных «подходящих» и «неподходящих» документов):
– если эффективность поиска растет, то группа признается приоритетной, и пользователь продолжает работу в рамках текущей модели (п. 5); дополнительно могут применяться другие решения из этой же группы (и тогда осуществляется переход к п. 4);
– в противном случае, если еще не все группы обработаны, то за основу выбирается очередная группа с меньшей степенью принадлежности пользователя к ней (и переходим к п. 3), иначе – на базе данного пользователя формируем новую группу и осуществляем мониторинг его дальнейших действий (п. 5).

Адаптивный поиск в корпоративных информационных системах

Применение адаптивного поиска в ИПС в среде Интернета сдерживается:
– длительностью процесса адаптации системы к индивидуальным особенностям конкретного пользователя – пока будет идти адаптация и подстройка поисковой системы, пользователь потеряет всякий интерес к дополнительным возможностям, которые он мог бы получить;
– естественной разноплановостью запросов одного и того же пользователя, которая не позволяет однозначно отнести его к одной определенной группе без наличия дополнительной учетной информации.

Другое дело, реализация адаптивного поиска в закрытой корпоративной информационной системе:
– всех пользователей можно сразу адекватно разбить на группы (например, по их профессиональным обязанностям, по продуктовым направлениям, продвижение которых они поддерживают и т.п.);
– пользователь сразу ощутит полезность адаптивного механизма (поиск уже будет подстроен под типичные потребности пользователей данной группы благодаря усилиям предыдущих менеджеров этой группы) и при реализации в компании адекватной корпоративной политики сам активно подключится к развитию и совершенствованию данного механизма.

При этом адаптивный поиск эффективен в любой из трех выделенных в работе [16, с. 24] моделей корпоративных информационных систем как механизм поиска:
– запасов ресурсов в «модели ресурсов», представляющей организацию как систему ресурсов (финансов, материальных запасов, кадров), принадлежащих юридическим или физическим лицам;
– документов и их реквизитов в «модели процессов», описывающих организацию как систему бизнес-процессов (управление проектами, документооборот, управление качеством и т.п.);
– знаний в «модели коллективов», рассматривающих организацию как систему небольших коллективов сотрудников, решающих общую задачу, при этом в роли организующих факторов выступают корпоративные знания и эффективные коммуникации.

Выводы

Подключение адаптивных механизмов к поисковым системам помогает пользователям найти нужную информацию за наиболее короткое время, делает сам поиск психологически приемлемым и адекватным. Применение предложенного в данной работе подхода к организации адаптивной поддержки сотрудничества позволит достичь синергии усилий пользователей, близких по своим информационно-поисковым запросам, при нахождении необходимой информации.

Поэтому особенно значимым на практике представляется использование данного подхода в рамках корпоративных информационных систем, где точность первоначальной информации о пользователе и о его принадлежности к имеющимся группам позволяет сразу достичь высокого уровня адаптированности поисковых возможностей системы к потребностям конкретного пользователя.

Предложенный подход может найти применение и при поиске информации в корпоративных СУБД и хранилищах данных. При этом поиск будет осуществляться среди метаданных, содержащих описания, например, нерегламентированных запросов к СУБД, первичных контролей входных данных, правил агрегирования данных или формирования матриц выходных таблиц.

Литература

1. ГОСТ 7.73-96. Поиск и распространение информации. Термины и определения [Текст]. – Взамен ГОСТ 7.27-80 ; введ. 1998-01-01. – Минск : Межгос. совет по стандартизации, метрологии и сертификации ; М. : Изд-во стандартов, 1997. – IV, 15 с. – (Система стандартов по информации, библиотечному и издательскому делу).
2. Капустин В.А. Протокол Z39.50 : обзор [Текст] / В.А. Капустин, В.М. Смирнов // Электронные библиотеки : перспективные методы и технологии, электронные коллекции : Первая Всероссийская научная конференция, 19 – 21 октября 1999 г. : труды конф. – СПб. : Изд-во СПб. университета, 1999. – С. 124-127. – Библиогр. : С. 126-127.
3. The Diverse and Exploding Digital Universe. An Updated Forecast of Worldwide Information Growth Through 2011 : An IDC White Paper [Text] / [Chute C., Manfrediz A., Minton S. et al.] ; J.F. Gantz (proj. dir.). – Framingham : IDC, March 2008. – 16 p. – Bibliogr. : P. 12-13.
4. My Yahoo! [Electronic resource]. – Режим доступа : http://my.yahoo.com/. – Загл. с экрана.
5. Customize Your Search Results with SearchWiki [Electronic resource]. – Режим доступа : http://www.google.com/intl/en/press/annc/20081120_searchwiki.html. – Загл. с экрана.
6. U Rank [Electronic resource]. – Режим доступа : http : //research.microsoft.com/en-us/projects/urank/. – Загл. с экрана.
7. Сороко В.Н. Функционально-структурное проектирование автоматизированных обучающих систем [Текст] / В.Н. Сороко, О.Р. Чертов // Інформатизація та нові технології. – 1995. – № 3-4. – С. 15-16. – Библиогр.: С. 16.

8. Brusilovsky P. Adaptive and Intelligent Web-based Educational Systems [Text] / P. Brusilovsky, Ch. Peylo // International Journal of Artificial Intelligence in Education. – 2003. – № 13. – P. 156-169. – Bibliogr. : P. 165-169.
9. Гагарін О.О. Дослідження і аналіз методів та моделей інтелектуальних систем безперервного навчання [Текст] / О.О. Гагарін, С.В. Титенко // Наукові вісті НТУУ «КПІ». – 2007. – № 6 (56). – С. 37-48. – Бібліогр. : С. 47-48.
10. Aubert H. Physiologie der Netzhaut [Text] / Hermann Aubert [Esther von Krosigk (Hrsg.)]. – Saarbrücken : Vdm Verlag Dr. Müller, 2007. – 412 p. – (Edition Classic). – ISBN 978-3836411530.
11. Скурихин В.И. Проектирование систем адаптивного управления производством [Текст] / В.И. Скурихин, В.А. Забродский, Ю.В. Копейченко. – Харьков : Вища школа : Изд-во при Харьк. ун-те, 1984. – 206 с. – Библиогр. : С. 200-204.
12. Акофф Р.О целеустремленных системах [Текст] / Р. Акофф, Ф. Эмери ; [перевод с англ. Г.Б. Рубальского ; под ред. И.А. Ушакова]. – М. : Советское радио, 1974. – 272 с.
13. Herring Ch. Adaptable and Adaptive Systems: The Intelligent Control Paradigm for Software Architecture [Text] / Charles Herring // Working Conference on Complex and Dynamic Systems Architectures : First Australian Working Conference, December 12 – 14, 2001 : Proceedings / Distributed Systems Technology Centre. – Brisbane, 2001. – ISBN 1864995823.
14. Райчук Д.В. Засоби адаптації інформаційно-пошукових систем [Текст] / Д.В. Райчук, О.Р. Чертов // Прикладна математика та комп'ютинг ПМК-2009 : (15 – 17 квітня 2009 р.) I наук. конф. / Нац. техн. ун-т України «Київський політехнічний інститут». – К. : НТУУ «КПІ», 2009. – С. 67-70. – Бібліогр. : С. 70. – ISBN 978-966-7115-69-2.
15. Sugiyama K. Adaptive web search based on user profile constructed without any effort from users [Text] / Kazunari Sugiyama, Kenji Hatano, Masatoshi Yoshikawa // International World Wide Web Conference : 13th international conference, May 17 – 22, 2004 : Proceedings / ACM. – NY, 2004. – P. 675-684. – Bibliogr. : P. 683-684. – ISBN 1-58113-844-X.
16. Кулькова Г.В. Электронные библиотеки корпоративных информационных систем [Текст] : обзор / Г.В. Кулькова, В.Н. Белоозеров, В.М. Ефременкова // Научно-техническая информация. Сер. 1, Организация и методика информационной работы. – 2006. – № 9. – С. 23-29. – Библиогр. : С. 29.

О.Р. Чертов, Д.В. Райчук
Адаптивна підтримка співробітництва під час пошуку інформації
В зв'язку зі значним зростанням обсягів інформації, доступної в цифровій формі, та збільшенням різноманіття веб-застосувань, необхідність в ефективному інформаційному пошуку стає критичною. В статті пропонується застосування технології адаптивної підтримки співробітництва, що забезпечить синергію роботи користувачів пошукової системи.

O.R. Chertov, D.V. Raichuk
Adaptive collaboration support during information retrieval
With the huge increase in the volume of information available in digital form and the increasing diversity of Web applications, the need for efficient information retrieval is critical. The application of adaptive collaboration support technology is proposed for synergy of search system users.

Статья поступила в редакцию 29.05.2009.